NUWC-NPT Technical Report 11,221
1 May 2000

Tracking on Intensity-Modulated Data Streams

Roy L. Streit
Submarine Combat Systems Directorate

Naval Undersea Warfare Center Division
Newport, Rhode Island

PREFACE

This report was prepared under NUWC Division Newport Project No. A196100, "Passive Sonar Advanced Processing Build," Program Element 63504N, principal investigator Walter R. Lane (Code 2123). The sponsoring activity is the Naval Sea Systems Command (SEA 93), Advanced Systems and Technology Office (ASTO, Robert E. Zarnich).

The technical reviewer for this report was Marcus L. Graham (Code 2211).

The author is grateful to Tod E. Luginbuhl (Code 2121) for his helpful technical discussions related to this report, to Marcus L. Graham (Code 2211) for his assistance in clarifying the data-dependent Bayesian density, and to Michael L. Walsh (Code 2211) for his careful reading of the entire manuscript, especially the section on the recursive algorithm.

Reviewed and Approved: 1 May 2000

John R. Short
Director, SubmarineCombat Systems

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	1 May 2000	

4. TITLE AND SUBTITLE

Tracking on Intensity-Modulated Data Streams

5. FUNDING NUMBERS

6. AUTHOR(S)

Roy L. Streit

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Naval Undersea Warfare Center Division
1176 Howell Street
Newport, RI 02841-1708

8. PERFORMING ORGANIZATION REPORT NUMBER

TR 11,221

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Naval Sea Systems Command (SEA 93, ASTO)
2531 Jefferson Davis Highway
Arlington, VA 22242-5160

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

The theoretical framework of probabilistic multi-hypothesis tracking is used to derive an algorithm for tracking one or more targets moving against a noisy background in an intensity-modulated, nonstationary data stream. The resulting multi-target tracking algorithm utilizes all the data available to a sensor display, and it completely avoids the current widespread practice of thresholding sensor data to generate point measurements that are subsequently fed to a tracker. The fundamental premise of the report is that tracking losses due to sensor data thresholding can be eliminated if the entire sensor output data set is utilized by the tracking algorithm.

14. SUBJECT TERMS

Target Tracking Target Tracking Algorithms Probabilistic Multi-Hypothesis Tracking

15. NUMBER OF PAGES
39

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	SAR

TABLE OF CONTENTS

TRACKING ON INTENSITY–MODULATED DATA STREAMS

1. INTRODUCTION

The theoretical methodology of probabilistic multi-hypothesis tracking (PMHT) [1] is extended to tracking one or more targets moving against a noisy background in an intensity-modulated, nonstationary data stream. The algorithm is called herein the histogram-PMHT algorithm because its derivation depends in a crucial way on a synthetic histogram interpretation of the measured data. The goal is to derive a multi-target tracking algorithm that utilizes all the data available to the sensor display, thereby completely avoiding the current widespread practice of thresholding the sensor data (e.g., peak picking) to generate point "measurements" that are subsequently fed to a tracker. Preprocessing the sensor data in this way results in a loss of information that may significantly increase the number of lost tracks and the tracking error in some applications. The fundamental premise of this report is that thresholding loss can be eliminated if the entire sensor output data set is properly utilized by the tracking algorithm. The histogram-PMHT algorithm derived in this report uses all the sensor output data.

Nonstationary, intensity-modulated data streams arise in many applications. A commonplace one-dimensional application is a waterfall (or similar) display of short-term, magnitude-squared, Fourier transform data versus frequency. For passive directional arrays, another one-dimensional application is a waterfall display of received broadband power as a function of bearing. A two-dimensional application to directional arrays is the sequence of intensity-modulated bearing-frequency images that arises if short-term spectra are computed on the arrays' directional beam outputs.

The development in this report is general in that sensor data of any dimensionality are treated, and the general version of PMHT [1] is utilized. It is shown that the histogram-PMHT algorithm is an iteratively reweighted Kalman filter for traditional single-scan updates. For batch data, the histogram-PMHT algorithm is an iteratively reweighted Kalman smoothing filter. Because histogram-PMHT uses a richer data set than PMHT, it requires more overhead (i.e., computer resources) than PMHT to compute the iteratively re-estimated synthetic measurements used in the algorithm; however, the computational complexity of the histogram-PMHT algorithm is otherwise equivalent to that of the PMHT algorithm.

An important theoretical aspect of the work is the unorthodox way in which a Bayesian *a priori* density on target state is added to the model. The usual approach fails in this problem because an intermediate measurement quantization procedure creates an infinite amount of synthetic histogram data (in the limit as the quantization level goes to zero), and this quantity of data overwhelms the usual prior density. The "trick" used here is to apply an *a priori* density to each point in the synthetically generated histogram, so that

the overall *a priori* density is properly "balanced" against the histogram data at every level of quantization. Using this data-dependent Bayesian density, the quantization step can be completely eliminated via a limiting argument. Consequently, the histogram-PMHT algorithm uses intensity-modulated sensor data and not synthetic histogram data.

Luginbuhl was the first to suggest the joint application of histogram and mixture modeling methodologies to intensity-modulated displays. In his Ph. D. thesis [2], he applied them to the problem of estimating a general, discrete-time, frequency-modulated process using a one-dimensional waterfall display of unaveraged, short-term, Fourier transform data. His methodology draws on several sources of earlier work. The fundamental idea of applying the method of expectation-maximization (EM) to histogram data is due to Dempster, Laird, and Rubin in their original EM paper [3]. Dempster and Rubin [4] used the EM method to derive Shepard's corrections for one-dimensional histogram data. McLachlan and Jones [5] and Jones and McLachlan [6,7] were the first to apply the EM method to estimate Gaussian mixtures from one-dimensional histogram data. Among these references [2-7], only Luginbuhl considered the important problem of linking successive histograms using a dynamical model.

This report is organized around the discussion of the fixed batch length problem, that is, when all the data from the initial to the current scan are processed jointly by the histogram-PMHT algorithm. The recursive problem, in which the batch is a fixed length time window that slides over a data stream, is discussed only after the intricate theoretical details are ironed out for the fixed batch problem. Throughout the report specific parameterizations are not introduced until they are needed. This approach not only minimizes the rather extensive notational difficulties that are common to algorithms developed by the method of EM, but it also clarifies the theoretical development.

Section 2 formulates the incomplete data likelihood function using the synthetic histogram point of view. Section 3 formulates the complete data likelihood function using three progressive stages of missing data. Section 4 evaluates the E-step, that is, the conditional expectation of the logarithm of the complete data likelihood function with respect to the missing data. This section discusses the data-dependent *a priori* density, and it also shows how to eliminate the synthetic histogram by taking the limit as the quantization level goes to zero. Section 5 gives the auxiliary function in a separable form suited to the multi-target application. Section 6 evaluates the M-step; that is, it solves for parameters that maximize the auxiliary function evaluated in the E-step. Section 7 is especially important because it states the histogram-PMHT algorithm for the fixed and recursive batch problems in a form suitable for implementation. Section 8 discusses modifications that can be incorporated into the algorithm and that may have utility for some applications. A summary is given in section 9.

2. INCOMPLETE DATA LIKELIHOOD FUNCTION

Let $C = \{C_1, ..., C_S\}$, $S \geq 1$, denote the collection of all possible sensor cells. It is assumed that $C_i \cap C_j = \phi$ for all i and j and that $C_1 \cup ... \cup C_S = R^{\dim(C)}$, where $\dim(C)$ denotes the dimension of the sensor space. The shape of a sensor cell may be very general; in fact, it need not even be connected (e.g., a cell may be the union of disjoint subintervals). The cells C are intrinsically fixed; however, those cells in which measurements are collected and displayed may vary from scan to scan. The sensor display at time t is denoted $B(t) = \{B_1(t), ..., B_{L(t)}(t)\} \subset C$, where $1 \leq L(t) \leq S$. The other measurement cells $B^c(t) = \{B_{L(t)+1}(t), ..., B_S(t)\} = C \setminus B(t)$ are not displayed and are said to be truncated. It is assumed that no measurement data are collected for cells in $B^c(t)$.

Let $T \geq 1$ denote the number of scans in a batch of measurements. The usual formulation for a recursive filter corresponds to the special case $T = 1$. Denote the sensor measurement vector at time t by

$$Z_t = \{z_{t1}, ..., z_{t,L(t)}\}, \qquad t = 1, ..., T, \tag{1}$$

where $z_{t\ell}$ is the magnitude-squared output of the sensor at time t in the displayed cell $B_\ell(t)$. Two kinds of variation must be modeled, one systematic and the other statistical. The systematic cell-to-cell variation of the measurement vector Z_t is parameterized using the PMHT multi-target model; that is, the unknown cell-to-cell variation is assumed to be proportional to a mixture density. In the PMHT approach, a target is either a single component in the mixture or a group of appropriately coupled components. A model of the statistical variation, or probability density function (PDF), of the individual measurement $z_{t\ell}$ must also be provided. The procedure adopted here is to quantize the data vector Z_t into a "pseudo-histogram," and then use a multinomial distribution to model the cell counts in the histogram. The expected cell counts are derived via the PMHT target mixture model. The appropriateness of these models must be verified on a case-by-case basis in the application.

Let $\hbar^2 > 0$ be a specified quantization level, and let

$$N_t = \{n_{t1}, ..., n_{t,L(t)}\}, \qquad t = 1, ..., T, \tag{2}$$

denote the quantized vector corresponding to Z_t, where

$$n_{t\ell} = \left\lfloor \frac{z_{t\ell}}{\hbar^2} \right\rfloor \tag{3}$$

and $\lfloor x \rfloor$ denotes the greatest integer less than or equal to x. The use of the quantized vector N_t instead of the measurement vector Z_t is an intermediate step in the development. In the sequel, after deriving the auxiliary function of the histogram-PMHT algorithm using the quantized vectors N_t, the measurement vectors Z_t will be recovered in the limit as $\hbar^2 \to 0$. Let

$$N_{t\Sigma} = \sum_{\ell=1}^{L(t)} n_{t\ell} \tag{4}$$

denote the total count, or sample size, at time t.

It is assumed that the vector N_t has a multinomial distribution consisting of $N_{t\Sigma}$ independent draws (with replacement) on $L(t)$ "categories" with probabilities

$$\frac{P_\ell(X_t)}{P(X_t)}, \qquad \ell = 1, ..., L(t), \tag{5}$$

where, for all cells,

$$P_\ell(X_t) = \int_{B_\ell(t)} f(\tau; X_t)\, d\tau, \qquad \ell = 1, ..., S, \tag{6}$$

and

$$P(X_t) = \sum_{\ell=1}^{L(t)} P_\ell(X_t), \tag{7}$$

where $f(\tau; X_t)$ denotes a "sample" PDF defined over all $\tau \in R^{\dim(C)}$, and where the vector X_t denotes the parameter vector of the sample PDF at time t. At the moment, X_t is not a random variable. In the sequel, $f(\tau; X_t)$ will be taken to be a Gaussian mixture in which, just as in PMHT, the mixture components correspond to targets.

The assumption of a multinomial density for the quantized data vector N_t is a nontrivial assumption. It is equivalent to the statement that the sample counts $N_t = \{n_{t1}, ..., n_{t,L(t)}\}$ form a histogram with cells $B_1(t), ..., B_{L(t)}(t)$ with a sample size of $N_{t\Sigma}$, where the samples are independent and identically distributed (IID) with the PDF $f(\tau; X_t)/P(X_t)$. Because N_t is generated via a synthetically imposed quantization procedure, the samples are not IID. The histogram model assumes, therefore, that more information is available than the measured data Z_t can possibly possess. The synthetic histogram data also cause difficulty when a Bayesian model for the parameters X_t is adopted because the abundance of synthetically generated, but supposedly IID, samples overwhelms the usual prior density. This mismatch is solved (see section 3.3 below) by choosing a sufficiently uninformative prior density to compensate for using a too informative likelihood function for the synthetic histogram data (i.e., the multinomial PDF).

Let $N = \{N_1, ..., N_T\}$ denote the collection of quantized measurement vectors, and let $X = \{X_1, ..., X_T\}$. Then, assuming that the vectors making up N are independent, the so-called incomplete data PDF of N is given by

$$p_{inc}(N; X) = \prod_{t=1}^{T} p_{inc}(N_t; X_t), \tag{8}$$

where

$$p_{inc}(N_t; X_t) = \frac{N_{t\Sigma}!}{n_{t,1}! \cdots n_{t,L(t)}!} \prod_{\ell=1}^{L(t)} \left[\frac{P_\ell(X_t)}{P(X_t)}\right]^{n_{t\ell}}. \tag{9}$$

The incomplete data likelihood function of X is obtained from (8) by substituting an appropriate, application-dependent, parametric form for the sample density $f(\tau; X_t)$ into (9).

If $p_\Xi(X)$ denotes the *a priori* density of X, then the incomplete data likelihood function is given by

$$p_{inc}(N, X) = p_\Xi(X)\ p_{inc}(N|X), \tag{10}$$

where the density $p_{inc}(N|X)$ is essentially identical to (8), the only difference being its statistical interpretation. A Bayesian prior density for the parameter X is developed in the next section (*cf.* equation (29)), and it must be included in the incomplete data density when the Bayesian viewpoint is adopted.

3. COMPLETE DATA LIKELIHOOD FUNCTION

3.1 UNOBSERVED CELL COUNTS AS MISSING DATA

Missing data are introduced in three progressive stages, with the Bayesian prior density on the parameters X being introduced at the end of the second stage. A suitable version of the general treatment of Dempster, Laird, and Rubin [3, Section 4.2] is followed initially; later, after mixture models are incorporated into the problem, the more detailed structure of McLachlan and Jones [5] is followed. For a general, up-to-date discussion of the method of EM and its variants, see McLachlan and Krishnan [8].

In the first stage, missing random variables are used to model the counts in the unobserved, or truncated, cells in $B^c(t)$. For $\ell = L(t)+1, ..., S$, let $n_{t\ell}$ denote the missing count for cell $B_\ell(t)$. It is assumed that the missing counts are distributed as a negative multinomial. (See Johnson et al. [9, chapter 36] for a general description of the negative multinomial and related discrete PDFs.) Letting

$$N_t^c = \{n_{t,L(t)+1}, \cdots, n_{t,S}\}, \tag{11}$$

and

$$N_{t\Sigma}^c = \sum_{\ell=L(t)+1}^{S} n_{t\ell}, \tag{12}$$

the negative multinomial PDF on N_t^c is given by

$$p(N_t^c|N_t; X_t) = \frac{(N_{t\Sigma} + N_{t\Sigma}^c - 1)!}{n_{t,L(t)+1}! \cdots n_{t,S}! \, (N_{t\Sigma} - 1)!} [P(X_t)]^{N_{t\Sigma}} \prod_{\ell=L(t)+1}^{S} [P_\ell(X_t)]^{n_{t\ell}}. \tag{13}$$

It is sufficient in (13) to condition only on the total count $N_{t\Sigma}$ instead of N_t; however, this notation is convenient for the present purpose. Because $p(N_t^c|N_t; X_t)$ is a PDF,

$$\sum_{N_t^c} p(N_t^c|N_t; X_t) = \sum_{n_{t,L(t)+1}=0}^{\infty} \cdots \sum_{n_{t,S}=0}^{\infty} p(N_t^c|N_t; X_t) = 1. \tag{14}$$

The conditional mean value of the missing cell count $n_{t\ell}$ is given by

$$\begin{aligned} \sum_{N_t^c} n_{t\ell} \, p(N_t^c|N_t; X_t) &= \sum_{n_{t,L(t)+1}=0}^{\infty} \cdots \sum_{n_{t\ell}=0}^{\infty} \cdots \sum_{n_{t,S}=0}^{\infty} n_{t\ell} \, p(N_t^c|N_t; X_t) \\ &= N_{t\Sigma} \frac{P_\ell(X_t)}{P(X_t)}, \qquad L(t)+1 \le \ell \le S, \end{aligned} \tag{15}$$

an intuitively reasonable result. Expression (15) is derived by substituting (13), canceling $n_{t\ell}$ in the $n_{t\ell}!$ term, and adjusting the remaining parameters in a straightforward way so that definition (13) can be re-invoked.

A model of the negative multinomial distribution in terms of realizations of a (stationary) Bernoulli trial sequence may provide an insightful and physically meaningful interpretation for some applications. Each trial in the sequence has S possible outcomes, one corresponding to each cell in C. As the sequence progresses, a running count of the total number of occurrences of each outcome is recorded, and the sequence stops when a total of $N_{t\Sigma}$ outcomes is obtained in the set of observed cells $\{B_1(t), ..., B_{L(t)}(t)\}$. Upon stopping the trial sequence, counts $N_t^c = \{n_{t,L(t)+1}, ..., n_{t,S}\}$ have been obtained in the unobserved cells $B_{L(t)+1}(t), ..., B_S(t)$. The probability of these counts is given by (13).

Let $N^c = \{N_1^c, ..., N_T^c\}$ denote the collection of missing measurement count vectors. Using independence of the count vectors in N^c and Bayes Theorem gives the complete data PDF at the end of the first stage as

$$\begin{aligned} p_{com}^{(1)}(N, N^c; X) &= \prod_{t=1}^{T} p_{com}^{(1)}(N_t, N_t^c; X_t) \\ &= \prod_{t=1}^{T} p_{inc}(N_t; X_t)\ p(N_t^c | N_t; X_t). \end{aligned}$$

Substituting (9) and (13), using definitions (4) and (12), and simplifying the resulting expression gives

$$p_{com}^{(1)}(N, N^c; X) = \prod_{t=1}^{T} \gamma_t \prod_{\ell=1}^{S} [P_\ell(X_t)]^{n_{t\ell}}, \tag{16}$$

where

$$\gamma_t = \frac{N_{t\Sigma}!}{n_{t,1}! \cdots n_{t,L(t)}!} \times \frac{(N_{t\Sigma} + N_{t\Sigma}^c - 1)!}{n_{t,L(t)+1}! \cdots n_{t,S}!\,(N_{t\Sigma} - 1)!}. \tag{17}$$

It is clear from (16) that the negative multinomial cancels denominator terms of the form $P(X_t)$ in the multinomial PDF.

3.2 SAMPLE LOCATIONS AS MISSING DATA

In the second stage, missing random variables are used to represent the location of the unobserved samples in all S cells. There are $n_{t\ell}$ samples at time t in cell $B_\ell(t)$, so let

$$\zeta_{t\ell} = \{\zeta_{t\ell 1}, ..., \zeta_{t\ell n_{t\ell}}\} \subset B_\ell(t) \tag{18}$$

denote the locations of the samples within cell $B_\ell(t)$. The random variables in $\zeta_{t\ell}$ are assumed to be IID with PDF $f(z; X_t)/P_\ell(X_t)$, and their domain is restricted to $B_\ell(t)$. Let

$$\zeta_t = \{\zeta_{t1}, ..., \zeta_{tS}\},$$

and

$$\zeta = \{\zeta_1, ..., \zeta_T\}.$$

The complete data PDF for the second stage is defined to be

$$p^{(2)}_{com}(N, N^c, \zeta; X) = \prod_{t=1}^{T} \gamma_t \prod_{\ell=1}^{S} \prod_{r=1}^{n_{t\ell}} f(\zeta_{t\ell r}; X_t). \tag{19}$$

The marginal density of (19) over ζ is easily seen to be identical to (16), so definition (19) is compatible with the existing statistical structure.

3.3 DATA-DEPENDENT BAYESIAN MODEL FOR X

Before continuing to the third stage of adding missing data, a Bayesian model is adopted for the parameter vector X. Following the usual approach, let

$$\Xi = \{\Xi_0, \Xi_1, ..., \Xi_T\}$$

denote the sequence of multi-target state vectors, or random variables, and let

$$X = \{X_0, X_1, ..., X_T\}$$

denote a realization of the state sequence Ξ. The additional state vector Ξ_0 is used to model the *a priori* variable of Ξ_1. Under a Markov assumption for Ξ, Bayes Theorem gives

$$p_\Xi(X) = p_{\Xi_0}(X_0) \prod_{t=1}^{T} p_{\Xi_t|\Xi_{t-1}}(X_t|X_{t-1}), \tag{20}$$

where $p_{\Xi_0}(X_0)$ is the density of Ξ_0. Applied to the complete data PDF (19), this approach yields the joint density

$$\tilde{p}^{(2)}_{com}(X, N, N^c, \zeta) = p_\Xi(X)\ p^{(2)}_{com}(N, N^c, \zeta|X). \tag{21}$$

The conditional density in (21) is obtained from (19) in the usual way by interpreting parametric dependence as Bayesian conditioning.

Formulation (21) has the surprising consequence that the *a priori* density $p_\Xi(X)$ has no influence on the parameter estimates as $N_{t\Sigma} \to \infty$ or, equivalently, as the arbitrary quantization scale factor $\hbar^2 \to 0$. The reason is that data counts in the histogram become infinite as $\hbar^2 \to 0$ and, consequently, the synthetically induced abundance of data overwhelms the *a priori* density. Since the information in the original sensor measurement vectors Z_t is independent of the degree of quantization, the traditional approach is not the appropriate way to pose a Bayesian model for the complete data PDF (19).

An alternative Bayesian formulation is adopted here. In this approach, the Bayesian prior density is applied to each event that generates a count in a cell. The pairs in the set

$$\Omega_t = \cup_{\ell=1}^{S} \{(\xi_{t\ell r}, \zeta_{t\ell r}) : \quad r = 1, ..., n_{t\ell}\} \tag{22}$$

are assumed to be IID samples of a joint PDF conditioned on the state realization $\Xi_{t-1} = X_{t-1}$, denoted by $p_{\Xi_t \Upsilon_t | \Xi_{t-1}}(\cdot, \cdot | X_{t-1})$, where Υ_t is the location random variable with sample value $\zeta_{t\ell r}$, and Ξ_t is the state random variable with sample value $\xi_{t\ell r}$. Applying Bayes Theorem and assuming that Υ_t is independent of Ξ_{t-1} when conditioned on Ξ_t, gives

$$\begin{aligned} p_{\Xi_t \Upsilon_t | \Xi_{t-1}} &= p_{\Xi_t | \Xi_{t-1}} \, p_{\Upsilon_t | \Xi_t \Xi_{t-1}} \\ &= p_{\Xi_t | \Xi_{t-1}} \, p_{\Upsilon_t | \Xi_t} \, . \end{aligned} \tag{23}$$

Conditional independence of sample pairs in Ω_t implies that the likelihood function of Ω_t is, using (23),

$$\begin{aligned} & \prod_{\ell=1}^{S} \prod_{r=1}^{n_{t\ell}} p_{\Xi_t \Upsilon_t | \Xi_{t-1}}(\xi_{t\ell r}, \zeta_{t\ell r} | X_{t-1}) \\ = & \prod_{\ell=1}^{S} \prod_{r=1}^{n_{t\ell}} p_{\Xi_t | \Xi_{t-1}}(\xi_{t\ell r} | X_{t-1}) \, p_{\Upsilon_t | \Xi_t}(\zeta_{t\ell r} | \xi_{t\ell r}). \end{aligned} \tag{24}$$

The Bayesian assumption, as it is applied in (24), is that each measurement $\zeta_{t\ell r}$ is generated from a parameter specific to it, namely $\xi_{t\ell r}$; thus, the total number of parameters $\{\xi_{t\ell r}\}$ equals, or balances, the total number of data points.

To make maximum *a posteriori* (MAP) estimation tractable and, more importantly, to make the data-specific parameters $\{\xi_{t\ell r}\}$ consistent with the parameter sequence $X = \{X_0, X_1, ..., X_T\}$, the parameters $\{\xi_{t\ell r}\}$ are constrained so that

$$\xi_{t\ell r} = X_t, \qquad \text{for } r = 1, ..., n_{t\ell}; \quad \ell = 1, ..., S. \tag{25}$$

The constraints (25) apply to realizations of the Bayesian random variable models of the parameters, not to the random variables themselves. Substituting the constraints (25) into (24) gives the likelihood function of the data Ω_t in the form

$$\left[p_{\Xi_t | \Xi_{t-1}}(X_t | X_{t-1}) \right]^{N_{t\Sigma} + N^c_{t\Sigma}} \prod_{\ell=1}^{S} \prod_{r=1}^{n_{t\ell}} p_{\Upsilon_t | \Xi_t}(\zeta_{t\ell r} | X_t). \tag{26}$$

Finally, the Bayesian version of (19) used here is given by

$$p^{(2)}_{com}(X, N, N^c, \zeta) = p_\Xi(X) \prod_{t=1}^{T} \gamma_t \prod_{\ell=1}^{S} \prod_{r=1}^{n_{t\ell}} f(\zeta_{t\ell r} | X_t), \tag{27}$$

where the obvious identification

$$p_{\Upsilon_t|\Xi_t}(\zeta_{t\ell r}|X_t) \equiv f(\zeta_{t\ell r}|X_t) \tag{28}$$

has been substituted into (27), and where the data-dependent prior density is

$$p_{\Xi}(X) = p_{\Xi_0}(X_0) \prod_{t=1}^{T} \left[p_{\Xi_t|\Xi_{t-1}}(X_t|X_{t-1})\right]^{N_{t\Sigma}+N^c_{t\Sigma}}. \tag{29}$$

The exponent of the density $p_{\Xi_0}(X_0)$ in (29) is unity, given the absence of synthetic histogram data at time $t = 0$. It will be seen that $p_{\Xi_0}(X_0)$ is eliminated in the limit as the quantization scale $\hbar^2 \to 0$.

The alternative density (29) has as many *a priori* density factors per scan as there are location samples in the synthetic histogram of that scan, in contrast to the usual density (20), which has only one such factor per scan. Hence, the data cannot overwhelm the *a priori* density (29) as the quantization scale $\hbar^2 \to 0$.

3.4 MIXTURE COMPONENT ASSIGNMENTS AS MISSING DATA

The final stage of missing random variables is required by the particular sample PDF considered in this application. The sample PDF is a function of location in the sensor output space $R^{\dim(C)}$, and it is assumed to be the mixture density

$$f(\tau|X_t) = \sum_{k=0}^{M} \pi_{tk}\ G_k(\tau|X_t), \tag{30}$$

where the mixing proportions $\pi_{tk} \geq 0$ and

$$\pi_{t0} + \pi_{t1} + ... + \pi_{tM} = 1,$$

and where $G_k(\tau|X_t)$ is a PDF for all k; i.e., it is nonnegative and its integral over τ is equal to 1 for all X_t. Specific parametric forms and different subsets of parameters in X_t will eventually be identified for each of the $M+1$ components in (30), but doing so at this point in the derivation needlessly obscures the discussion.

A physical interpretation of component $\pi_{t0}G_0(\tau|X_t)$ is that π_{t0} represents the fraction of the total power due to the background and $G_0(\tau|X_t)$ models the variation from cell to cell of the background level. The remaining components $\pi_{t1}G_1(\tau|X_t), ..., \pi_{tM}G_M(\tau|X_t)$ are interpreted as target models in which π_{tk} is the fraction of total power due to target k and $G_k(\tau|X_t)$ models the cell-level variations of target k. The parametric form (30) assumes that a target's power level may be spread across more than one cell of the sensor display. The parametric form of the spread functions $G_k(\tau|X_t)$ need not be specified until later.

A missing variable $k_{t\ell r}$ is used to specify which component of the mixture generated the missing variable $\zeta_{t\ell r}$, so that $0 \leq k_{t\ell r} \leq M$. It is assumed that $k_{t\ell r}$ is a random variable with discrete PDF specified by $\{\pi_{t0}, \pi_{t1}, ..., \pi_{tM}\}$. Hence, if

$$K_{t\ell} = \{k_{t\ell 1}, ..., k_{t\ell n_{t\ell}}\} \qquad \text{for } \ell = 1, ..., S \text{ and } t = 1, ..., T,$$

then all variables in $K_{t\ell}$ are IID. Let $K_t = \{K_{t1}, ..., K_{tS}\}$ and $K = \{K_1, ..., K_T\}$. Extending the density (27) to include K gives, finally, the complete data PDF at the end of the third stage as

$$p^{(3)}_{com}(X, N, N^c, \zeta, K) = p_\Xi(X) \prod_{t=1}^{T} \gamma_t \prod_{\ell=1}^{S} \prod_{r=1}^{n_{t\ell}} f_{k_{t\ell r}}(\zeta_{t\ell r}|X_t), \tag{31}$$

where

$$f_k(\tau|X_t) = \pi_{tk}\ G_k(\tau|X_t) \tag{32}$$

is used in (31). The dependence of $f_k(\tau|X_t)$ on the mixing proportion π_{tk} is implicit in the abbreviated notation (32). Summing (31) over all K gives (27) after substitution of the mixture (30), so the additional random variables K are compatible with the existing structure.

4. E-STEP

4.1 CONDITIONAL EXPECTATION OVER ASSIGNMENTS K

In the E-step of the EM method, the so-called auxiliary function $Q_\hbar$ is evaluated as a conditional expectation of the logarithm of the complete data density (31). The required expectation is with respect to the missing data $\{N^c, \zeta, K\}$, and it is conditioned on N and a current value of X, denoted by X'. Explicitly,

$$Q_\hbar = E_{N^c \zeta K}\left[\{\log p^{(3)}_{com}(X, N, N^c, \zeta, K)\} \mid N, X'\right], \tag{33}$$

where $E_{N^c \zeta K}$ denotes the expectation with respect to the missing data. The conditional density of the missing data is obtained by using Bayes Theorem, the complete data density (31), and the incomplete data density (10). After algebraic manipulation, the result is

$$\begin{aligned} p(N^c, \zeta, K | N, X') &= \frac{p^{(3)}_{com}(X', N, N^c, \zeta, K)}{p_{inc}(N, X')} \\ &= \prod_{t=1}^{T} p(N^c_t | N_t, X'_t) \prod_{\ell=1}^{S} \prod_{r=1}^{n_{t\ell}} \frac{f_{k_{t\ell r}}(\zeta_{t\ell r} | X'_t)}{P_\ell(X'_t)}, \end{aligned} \tag{34}$$

where $p(N^c_t | N_t, X')$ is the negative multinomial density (13). Substituting (29) and (31) into (33), and dropping the terms γ_t because they do not depend on X, gives

$$\begin{aligned} Q_\hbar &= E_{N^c}\left[\log p_{\Xi_0}(X_0) + \sum_{t=1}^{T} N^{total}_{t\Sigma} \log p_{\Xi_t | \Xi_{t-1}}(X_t | X_{t-1}) \,\middle|\, N, X'\right] \\ &\quad + E_{N^c \zeta K}\left[\sum_{t=1}^{T}\sum_{\ell=1}^{S}\sum_{r=1}^{n_{t\ell}} \log f_{k_{t\ell r}}(\zeta_{t\ell r} | X_t) \,\middle|\, N, X'\right], \end{aligned} \tag{35}$$

where

$$N^{total}_{t\Sigma} = N_{t\Sigma} + N^c_{t\Sigma}, \qquad t = 1, ..., T.$$

The second expectation in (35) is evaluated as three nested, conditional expectations:

$$Q_\hbar = E_{N^c}\left[\, E_\zeta\left[\, E_K\left[\, \{\cdot\} \mid N, X'\right] \,\right] \,\right]. \tag{36}$$

The outermost expectation in (36) and the first expectation in (35) are both with respect to N^c.

The innermost expectation in (36) is, by definition,

$$\begin{aligned} E_K &= E_K\left[\sum_{t=1}^{T}\sum_{\ell=1}^{S}\sum_{r=1}^{n_{t\ell}} \log f_{k_{t\ell r}}(\zeta_{t\ell r} | X'_t) \,\middle|\, N, X'\right] \\ &= \sum_{K}\left\{\sum_{t=1}^{T}\sum_{\ell=1}^{S}\sum_{r=1}^{n_{t\ell}} \log f_{k_{t\ell r}}(\zeta_{t\ell r} | X_t)\right\} p(N^c, \zeta, K | N, X'). \end{aligned}$$

Interchanging the sum over K with the triple sum over $\{t, \ell, r\}$, substituting (34), and then pushing the sum over $K \backslash k_{t\ell r}$ further inside the summand gives

$$\begin{aligned} E_K &= \sum_{t=1}^{T} \sum_{\ell=1}^{S} \sum_{r=1}^{n_{t\ell}} \sum_{K} \log f_{k_{t\ell r}}(\zeta_{t\ell r}|X_t)\ p(N^c, \zeta, K|N, X') \\ &= \sum_{t=1}^{T} \sum_{\ell=1}^{S} \sum_{r=1}^{n_{t\ell}} \left\{ \sum_{k_{t\ell r}=0}^{M} \frac{p(N_t^c|N_t, X_t')}{P_\ell(X_t')}\ f_{k_{t\ell r}}(\zeta_{t\ell r}|X_t')\ \log f_{k_{t\ell r}}(\zeta_{t\ell r}|X_t) \right. \\ &\qquad \left. \times \sum_{K \backslash k_{t\ell r}} p(N^c, \zeta, K|N, X') \right\}. \end{aligned} \tag{37}$$

It is straightforward to see from (34) that

$$\sum_{K \backslash k_{t\ell r}} p(N^c, \zeta, K|N, X') = \prod_{\substack{\tilde{t}=1 \\ \tilde{t} \neq t}}^{T} p(N_{\tilde{t}}^c|N_{\tilde{t}}, X_{\tilde{t}}') \prod_{\substack{\tilde{\ell}=1 \\ \tilde{\ell} \neq \ell}}^{S} \prod_{\tilde{r}=1}^{n_{t\tilde{\ell}}} \frac{f(\zeta_{\tilde{t}\tilde{\ell}\tilde{r}}|X_{\tilde{t}}')}{P_{\tilde{\ell}}(X_{\tilde{t}}')}. \tag{38}$$

The absence of a subscript on the sample density $f(\zeta_{\tilde{t}\tilde{\ell}\tilde{r}}|X_{\tilde{t}}')$ in (38) is a direct consequence of marginalizing over K. This completes the conditional expectation with respect to K.

4.2 CONDITIONAL EXPECTATION OVER LOCATIONS ζ

The conditional expectation of E_K with respect to the collection of all sample locations ζ is defined by the multiple integral

$$E_\zeta = \int \cdots \int E_K\ d\zeta, \tag{39}$$

where there are as many integrals in (39) as there are indices in K. For all indices $\{t, \ell, r\}$, the domain of integration of $\zeta_{t\ell r}$ is $B_\ell(t)$. Substituting (37), pushing the multiple integral inside the triple sum over $\{t, \ell, r\}$, and then arranging the integrals to match corresponding sums over K gives

$$\begin{aligned} E_\zeta &= \sum_{t=1}^{T} \sum_{\ell=1}^{S} \sum_{r=1}^{n_{t\ell}} \left\{ \sum_{k_{t\ell r}=0}^{M} \frac{p(N_t^c|N_t, X_t')}{P_\ell(X_t')} \right. \\ &\qquad \times \int_{B_\ell(t)} f_{k_{t\ell r}}(\zeta_{t\ell r}|X_t')\ \log f_{k_{t\ell r}}(\zeta_{t\ell r}|X_t)\ d\zeta_{t\ell r} \\ &\qquad \left. \times \int \cdots \int_{\zeta \backslash \zeta_{t\ell r}} \sum_{K \backslash k_{t\ell r}} p(N^c, \zeta, K|N, X')\ d(\zeta \backslash \zeta_{t\ell r}). \right\} \end{aligned} \tag{40}$$

From (38) and the definition (30), it is seen that

$$\int_{\zeta\backslash\zeta_{t\ell r}} \cdots \int \sum_{K\backslash k_{t\ell r}} p(N^c, \zeta, K|N, X')\; d(\zeta\backslash\zeta_{t\ell r}) = \prod_{\substack{\tilde{t}=1 \\ \tilde{t}\neq t}}^{T} p(N^c_{\tilde{t}}|N_{\tilde{t}}, X'_{\tilde{t}}). \tag{41}$$

Expression (40) is simplified by performing the following steps: (1) substitute (41) into (40); (2) change $k_{t\ell r}$ and $\zeta_{t\ell r}$ to the dummy variables k and τ, respectively; (3) make the sum on k the outermost sum instead of the innermost; (4) collect terms to eliminate the exception in the product and recover the negative multinomial density; and (5) recognize that the summand no longer depends on r and so is equivalent to a multiplication. The result is

$$E_\zeta = \sum_{k=0}^{M} \sum_{t=1}^{T} \sum_{\ell=1}^{S} n_{t\ell} \frac{\int_{B_\ell(t)} f_k(\tau|X'_t)\; \log f_k(\tau|X_t)\; d\tau}{P_\ell(X'_t)} \prod_{\tilde{t}=1}^{T} p(N^c_{\tilde{t}}|N_{\tilde{t}}, X'_{\tilde{t}}). \tag{42}$$

This completes the conditional expectation with respect to ζ.

4.3 CONDITIONAL EXPECTATION OVER COUNTS N^c

The auxiliary function $Q_{\hbar}$ is, from (35), the sum of two terms, namely,

$$Q_{\hbar} = E_{N^c}\left[\log p_\Xi(X)|N, X'\right] + E_{N^c}\left[E_\zeta\right], \tag{43}$$

where the data-dependent *a priori* density $p_\Xi(X)$ is given by (29) and E_ζ is given by (42). The first term in (43) is a conditional expectation with respect to $N^c = \{N^c_1, ..., N^c_T\}$, where N^c_t is given by (11); however, the second term is an unconditional sum over all N^c because the required conditioning is implicit in the definition of E_ζ.

Both terms in (43) require evaluating the quantity $\overline{n}_{t\ell}$, where

$$\overline{n}_{t\ell} = E_{N^c}[n_{t\ell}|N, X'] = \sum_{N^c} n_{t\ell} \prod_{\tilde{t}=1}^{T} p(N^c_{\tilde{t}}|N_{\tilde{t}}, X'_{\tilde{t}}). \tag{44}$$

The expectation in (44) simplifies in one of two ways. If $1 \leq \ell \leq L(t)$, then $\overline{n}_{t\ell} = n_{t\ell}$ because $n_{t\ell}$ is not involved in the sum over N^c, and because

$$\sum_{N^c} \prod_{\tilde{t}=1}^{T} p(N^c_{\tilde{t}}|N_{\tilde{t}}, X'_{\tilde{t}}) = \left(\sum_{N^c_1} p(N^c_1|N_1, X'_1)\right) \cdots \left(\sum_{N^c_T} p(N^c_T|N_T, X'_T)\right), \tag{45}$$

evaluates to 1, using the PDF identity (14). If, however, $L(t) + 1 \leq \ell \leq S$, then

$$\overline{n}_{t\ell} = \sum_{N^c_t} n_{t\ell}\; p(N^c_t|N_t, X'_t) \left[\sum_{N^c\backslash N^c_t} \prod_{\substack{\tilde{t}=1 \\ \tilde{t}\neq t}}^{T} p(N^c_{\tilde{t}}|N_{\tilde{t}}, X'_{\tilde{t}})\right]. \tag{46}$$

The term in brackets in (46) equals 1, as is seen by factoring the sum as in (45); therefore, from (15),

$$\overline{n}_{t\ell} = \begin{cases} n_{t\ell} & \text{for } 1 \le \ell \le L(t) \\ N_{t\Sigma} \frac{P_\ell(X'_t)}{P(X'_t)} & \text{for } L(t)+1 \le \ell \le S. \end{cases} \tag{47}$$

Truncated cells are seen from (47) to contribute to $Q_\hbar$ in proportion to the expected number of measurements in those cells; thus, the negative multinomial PDF is a kind of extrapolation procedure to compensate for truncated cells.

By linearity of the expectation operator,

$$\begin{aligned} E_{N^c}\left[N_{t\Sigma}^{total} | N_t, X'_t\right] &= E_{N^c}\left[N_{t\Sigma} + N_{t\Sigma}^c | N_t, X'_t\right] \\ &= N_{t\Sigma} + \sum_{N^c} N_{t\Sigma}^c \prod_{\tilde{t}=1}^{T} p(N_{\tilde{t}}^c | N_{\tilde{t}}, X'_{\tilde{t}}) \\ &= N_{t\Sigma} + \sum_{\ell=L(t)+1}^{S} \overline{n}_{t\ell}. \end{aligned}$$

Substituting (47) and using the identity

$$P(X'_t) + \sum_{\ell=L(t)+1}^{S} P_\ell(X'_t) = \sum_{\ell=1}^{S} P_\ell(X'_t) = \int_{R^{\dim(C)}} f(\tau | X'_t)\, d\tau = 1$$

gives the result

$$E\left[N_{t\Sigma}^{total} | N_t, X'_t\right] = \frac{N_{t\Sigma}}{P(X'_t)}. \tag{48}$$

The results (47) and (48) enable the computation of (43).

The first term of (43) is evaluated from (35) using linearity of the conditional expectation operator and the limit (48) to obtain

$$E_{N^c}[\log p_\Xi(X) | N_t, X'_t] = \log p_{\Xi_0}(X_0) + \sum_{t=1}^{T} \frac{N_{t\Sigma}}{P(X'_t)} \log p_{\Xi_t | \Xi_{t-1}}(X_t | X_{t-1}). \tag{49}$$

The second term of (43) is evaluated from (42) using linearity of the expectation operator and substituting (47) to obtain

$$E_{N^c}[E_\zeta] = \sum_{k=0}^{M} \sum_{t=1}^{T} \sum_{\ell=1}^{S} \frac{\overline{n}_{t\ell}}{P_\ell(X'_t)} \int_{B_\ell(t)} f_k(\tau | X'_t) \log f_k(\tau | X_t)\, d\tau. \tag{50}$$

The auxiliary function $Q_\hbar$ is the sum of (49) and (50).

4.4 LIMIT OF $Q_\hbar$ AS QUANTIZATION SCALE $\hbar^2 \to 0$

It is now shown that the synthetic histogram is eliminated by taking the limit

$$Q^\sharp = \lim_{\hbar^2 \to 0} \hbar^2 \, Q_\hbar. \tag{51}$$

From the definition (3) of $n_{t\ell}$, it follows that

$$\lim_{\hbar^2 \to 0} \hbar^2 \, n_{t\ell} = \lim_{\hbar^2 \to 0} \hbar^2 \left\lfloor \frac{z_{t\ell}}{\hbar^2} \right\rfloor = z_{t\ell}. \tag{52}$$

From (3) and (4) it follows immediately that

$$\begin{aligned} \lim_{\hbar^2 \to 0} \hbar^2 \, N_{t\Sigma} &= \lim_{\hbar^2 \to 0} \left\{ \hbar^2 \sum_{\ell=1}^{L(t)} \left\lfloor \frac{z_{t\ell}}{\hbar^2} \right\rfloor \right\} \\ &= \sum_{\ell=1}^{L(t)} \lim_{\hbar^2 \to 0} \left\{ \hbar^2 \left\lfloor \frac{z_{t\ell}}{\hbar^2} \right\rfloor \right\} \\ &\equiv \|Z_t\|, \end{aligned} \tag{53}$$

where $\|\cdot\|$ denotes the so-called L_1-norm defined by

$$\|Z_t\| = \sum_{\ell=1}^{L(t)} z_{t\ell}. \tag{54}$$

If $z_{t\ell}$ is magnitude-squared data, then $\|Z_t\|$ is a sum of squares (i.e., it is proportional to power). Using (53), it is reasonable to define the expected measurement $\overline{z}_{t\ell}$ as

$$\begin{aligned} \overline{z}_{t\ell} &= \lim_{\hbar^2 \to 0} \hbar^2 \, \overline{n}_{t\ell} \\ &= \begin{cases} z_{t\ell} & \text{for } 1 \le \ell \le L(t) \\ \|Z_t\| \, \frac{P_\ell(X_t')}{P(X_t')} & \text{for } L(t)+1 \le \ell \le S. \end{cases} \end{aligned} \tag{55}$$

The limit (51) is evaluated term-by-term using expressions (52) and (55). The quantity $\hbar^2 \log p_{\Xi_0}(X_0)$ goes to 0 in the limit, and the final result is

$$\begin{aligned} Q^\sharp &= \sum_{t=1}^{T} \frac{\|Z_t\|}{P(X_t')} \log p_{\Xi_t|\Xi_{t-1}}(X_t|X_{t-1}) \\ &\quad + \sum_{k=0}^{M} \sum_{t=1}^{T} \sum_{\ell=1}^{S} \frac{\overline{z}_{t\ell}}{P_\ell(X_t')} \int_{B_\ell(t)} f_k(\tau|X_t') \, \log f_k(\tau|X_t) \, d\tau. \end{aligned} \tag{56}$$

This completes the evaluation of the limiting form of the auxiliary function. The limiting form $Q^\sharp$ uses measured sensor data, not the synthetic histogram data.

5. AUXILIARY FUNCTION FOR HISTOGRAM-PMHT

Target-specific subsets of the multi-target state and parameter vectors are now identified. The multi-target state vector Ξ_t is partitioned as

$$\Xi_t = \{\Xi_{t0}, \Xi_{t1}, ..., \Xi_{tM}\}, \qquad t = 0, 1, ..., T,$$

where Ξ_{tk} is the state of target k at time t. The corresponding target parameters are given by

$$X_t = \{x_{t0}, x_{t1}, ..., x_{tM}\}, \qquad t = 0, 1, ..., T,$$

where x_{t0} is the parameter of the background density at time t, and x_{tk} is the parameter of target k at time t. There is no need to specify a parametric form for $p_{\Xi_0}(X_0)$ because it is absent from $Q^\sharp$. Assuming that targets and background are independent at all times,

$$p_{\Xi_t|\Xi_{t-1}}(X_t|X_{t-1}) = \prod_{k=0}^{M} p_{\Xi_{tk}|\Xi_{t-1,k}}(x_{tk}|x_{t-1,k}). \tag{57}$$

The abbreviated notation in (32) rewritten in single target style becomes

$$f_k(\tau|X_t) = \pi_{tk}\, G_k(\tau|x_{tk}). \tag{58}$$

Mixing proportions and target parameters are to be estimated.

Let x'_{tk} and π'_{tk} denote current values of the target parameters and mixing proportions, respectively. Substituting (57) and (58) into (56) gives

$$Q^\sharp = \sum_{t=1}^{T} Q_{t\pi} + \sum_{k=0}^{M} Q_{kX}, \tag{59}$$

where

$$Q_{t\pi} = \sum_{k=0}^{M} \left[\sum_{\ell=1}^{S} \frac{\overline{z}_{t\ell}}{P_\ell(X'_t)} \int_{B_\ell(t)} G_k(\tau|x'_{tk})\, d\tau \right] \pi'_{tk} \log \pi_{tk} \tag{60}$$

and

$$\begin{aligned} Q_{kX} \;=\; & \sum_{t=1}^{T} \frac{\|Z_t\|}{P(X'_t)} \log p_{\Xi_t|\Xi_{t-1}}(x_{tk}|x_{t-1,k}) \\ & + \sum_{t=1}^{T} \sum_{\ell=1}^{S} \frac{\pi'_{tk}\, \overline{z}_{t\ell}}{P_\ell(X'_t)} \int_{B_\ell(t)} G_k(\tau|x'_{tk}) \log G_k(\tau|x_{tk})\, d\tau, \end{aligned} \tag{61}$$

where $\|Z_t\|$ and $\overline{z}_{t\ell}$ are given by (54) and (55), respectively.

The auxiliary function (59) holds for general parametric forms of the Markovian densities $p_{\Xi_t|\Xi_{t-1}}(x_{tk}|x_{t-1,k})$. It also holds for general parametric forms of both the target densities $G_k(\tau|x_{tk})$, $k \geq 1$, and the background density $G_0(\tau|x_{tk})$.

Different parameterizations will typically be used in most applications to model background structure and target behavior. For example, if the background model does not use a state process model, and it is stationary, so that $x_{t0} \equiv x_0$ for all t, then the contribution in (61) for the background term $k = 0$ becomes simply

$$Q_{0X} = \sum_{t=1}^{T} \sum_{\ell=1}^{S} \frac{\pi'_{t0}\ \overline{z}_{t\ell}}{P_\ell(X'_t)} \int_{B_\ell(t)} G_0(\tau|x'_0)\ \log G_0(\tau|x_0)\ d\tau. \tag{62}$$

Estimates of the stationary background parameter x_0 and the target parameters would then be derived from (62) and Q_{kX}, $1 \leq k \leq M$. Alternatively, if the background is not estimated because it is already normalized in some application-specific manner, then the term Q_{0X} is omitted from $Q^\sharp$ altogether.

6. M-STEP

6.1 MIXING PROPORTIONS: GENERAL CASE

An expression for updated mixing proportions is derived from the general expression (60). There is no need to specify particular parametric forms for the target and background densities $G_k(\tau|x'_{tk})$. The update, denoted by $\widehat{\pi}_{tk}$, is derived by maximizing (60) subject to the natural constraint on the sum of the π_{tk}'s. The appropriate Lagrangian function is

$$\sum_{k=0}^{M}\left[\sum_{\ell=1}^{S}\frac{\overline{z}_{t\ell}}{P_\ell(X'_t)}\int_{B_\ell(t)}G_k(\tau|x'_{tk})\,d\tau\right]\pi'_{tk}\log\pi_{tk}+\lambda_t\left(1-\sum_{k=0}^{M}\pi_{tk}\right),$$

where λ_t is the Lagrange multiplier. Differentiating with respect to π_{tk} and solving for π_{tk} gives the update

$$\widehat{\pi}_{tk}=\frac{\pi'_{tk}}{\lambda_t}\left[\sum_{\ell=1}^{S}\frac{\overline{z}_{t\ell}}{P_\ell(X'_t)}\int_{B_\ell(t)}G_k(\tau|x'_{tk})\,d\tau\right]. \tag{63}$$

Summing (63) over all k and invoking the equality constraint gives the Lagrange multiplier as

$$\lambda_t=\sum_{k=0}^{M}\pi'_{tk}\left[\sum_{\ell=1}^{S}\frac{\overline{z}_{t\ell}}{P_\ell(X'_t)}\int_{B_\ell(t)}G_k(\tau|x'_{tk})\,d\tau\right]. \tag{64}$$

Substituting (64) into (63) gives the updated estimate $\widehat{\pi}_{tk}$.

Additional linear constraints are easily imposed on the method. For example, it may be desirable to require the mixing proportions to be stationary; that is, $\pi_{tk}=\pi_k$ for all t. The resulting estimator requires summing over ℓ and t and adjusting the normalization factor accordingly. Such constraints are application-dependent and are not pursued here.

6.2 STATE ESTIMATES FOR LINEAR GAUSS-MARKOV CASE

Parameterizations appropriate for the special case of linear Gauss-Markov target processes, linear measurement models, and known background are assumed in this section. The measurement and target covariance matrices $\{R_{tk}\}$ and $\{Q_{tk}\}$ are assumed known throughout this section. The case of unknown covariance matrices is treated in the following section.

For $k=1,...,M$ and $t=1,...,T$, the target process models are

$$p_{\Xi_{tk}|\Xi_{t-1,k}}(x_{tk}|x_{t-1,k})=N(x_{tk};\ F_{t-1,k}\,x_{t-1,k},\ Q_{t-1,k}), \tag{65}$$

and the measurement models are

$$G_k(\tau|x_{tk})=N(\tau;\ H_{tk}\,x_{tk},\ R_{tk}), \tag{66}$$

where $N(\tau;\, \mu,\, \Sigma)$ denotes the multivariate Gaussian PDF with mean vector μ and covariance matrix Σ. The background density is assumed to be completely known for all t; that is,

$$G_0(\tau|x_{t0}) \equiv G_0(\tau; \overline{x}_{t0}), \tag{67}$$

where $\overline{x}_{t0}$ are known constants. Consequently, Q_{0X} is omitted from $Q^\sharp$.

For $k \geq 1$, substituting these forms into Q_{kX} and dropping determinant terms, since they do not depend on $\{x_{tk}\}$, shows that $-2Q_{kX}$ is a discrete-continuous sum of weighted, squared errors. Explicitly, the total squared error is given by

$$\begin{aligned} -2\, Q_{kX} \;=\; & \sum_{t=1}^{T} \frac{\|Z_t\|}{P(X_t')} \left(x_{tk} - F_{t-1,k} x_{t-1,k}\right)^* Q_{t-1,k}^{-1} \left(x_{tk} - F_{t-1,k} x_{t-1,k}\right) \\ & + \sum_{t=1}^{T} \sum_{\ell=1}^{S} \frac{\pi_{tk}'\, \overline{z}_{t\ell}}{P_\ell(X_t')} \int_{B_\ell(t)} N(\tau;\; H_{tk} x_{tk}',\; R_{tk}) \\ & \times \left(\tau - H_{tk} x_{tk}\right)^* R_{tk}^{-1} \left(\tau - H_{tk} x_{tk}\right)\, d\tau, \end{aligned} \tag{68}$$

where asterisks denote vector and matrix transpose. Let

$$X(k) = \{x_{0k}, x_{1k}, ..., x_{Tk}\}.$$

Using the general gradient identity

$$\nabla_x\; (Fx - \mu)^*\; \Sigma^{-1}\; (Fx - \mu) = 2\; F^*\; \Sigma^{-1}\; (Fx - \mu)$$

to take the gradient of the total squared error with respect to each of the vectors in $X(k)$ gives the necessary conditions for minimizing $-2Q_{kX}$. It is assumed that this linear system of equations is full rank, so that the necessary conditions are sufficient also in this case.

The linear system takes the form $\Gamma_k X(k) = b_k$, where Γ_k is a symmetric and block tridiagonal matrix of size $(T+1) \times (T+1)$ blocks and the right-hand side is a compatibly partitioned vector. After algebraic manipulation, the system matrix Γ_k is given explicitly as

$$\begin{bmatrix} D_{0k} & -B_{0k} & 0 & \cdots & 0 & 0 & 0 \\ -B_{0k}^* & A_{1k} + D_{1k} & -B_{1k} & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -B_{T-2,k}^* & A_{T-1,k} + D_{T-1,k} & -B_{T-1,k} \\ 0 & 0 & 0 & \cdots & 0 & -B_{T-1,k}^* & A_{Tk} \end{bmatrix},$$

where the block matrices A_{tk}, D_{tk}, and B_{tk} are given by

$$\begin{aligned} A_{tk} &= \left(\widetilde{Q}_{t-1,k}\right)^{-1} + H_{tk}^* \left(\widetilde{R}_{tk}\right)^{-1} H_{tk} && \text{for } 1 \leq t \leq T, \\ D_{tk} &= F_{tk}^* \left(\widetilde{Q}_{tk}\right)^{-1} F_{tk} && \text{for } 0 \leq t \leq T-1, \\ B_{tk} &= F_{tk}^* \left(\widetilde{Q}_{tk}\right)^{-1} && \text{for } 0 \leq t \leq T-1, \end{aligned} \tag{69}$$

and the synthetic target and measurement covariances used in (69) are

$$\widetilde{Q}_{tk} = \frac{P(X'_{t+1})}{\|Z_{t+1}\|}\, Q_{tk} \quad \text{for } 0 \le t \le T-1, \tag{70}$$

and

$$\widetilde{R}_{tk} = \frac{R_{tk}}{\pi'_{tk}\ \nu_{tk}} \quad \text{for } 1 \le t \le T, \tag{71}$$

where

$$\nu_{tk} = \sum_{\ell=1}^{S} \frac{\overline{z}_{t\ell}}{P_\ell(X'_t)} \int_{B_\ell(t)} N(\tau;\ H_{tk}x'_{tk},\ R_{tk})\, d\tau. \tag{72}$$

The right-hand side of the system is a vector that, partitioned into $T+1$ blocks, is given by

$$b_k = \begin{bmatrix} 0 \\ H^*_{1k}\ \left(\widetilde{R}_{1k}\right)^{-1}\ \widetilde{z}_{1k} \\ \vdots \\ H^*_{Tk}\ \left(\widetilde{R}_{Tk}\right)^{-1}\ \widetilde{z}_{Tk} \end{bmatrix}, \tag{73}$$

where the synthetic measurements used in (73) are given by

$$\widetilde{z}_{tk} = \frac{1}{\nu_{tk}} \sum_{\ell=1}^{S} \frac{\overline{z}_{t\ell}}{P_\ell(X'_t)} \int_{B_\ell(t)} \tau\ N(\tau;\ H_{tk}x'_{tk},\ R_{tk})\, d\tau. \tag{74}$$

The solution of the linear system $\Gamma_k X(k) = b_k$ gives updated state estimates, denoted by $\widehat{X}(k) = \{\widehat{x}_{0k}, \widehat{x}_{1k}, ..., \widehat{x}_{Tk}\}$. It can be solved using an appropriate form of Gaussian elimination for symmetric, block tridiagonal matrices.

Alternatively, expressions (69) and (73) are in precisely the form needed to show that a nonstationary Kalman smoothing filter with a diffuse prior PDF for the state at time $t = 1$ can also compute the same state estimates. This numerical equivalence is seen from the expression

$$\exp\left(Q_{kX}\right) \propto \prod_{t=1}^{T} N(x_{tk};\ F_{t-1,k}\, x_{t-1,k},\ \widetilde{Q}_{t-1,k})\ N(\widetilde{z}_{tk};\ H_{tk}\, x_{tk},\ \widetilde{R}_{tk}), \tag{75}$$

a result obtained from Q_{kX} by algebraically completing the squares on the variables in $X(k)$. Maximizing (75) over $X(k) = \{x_{0k}, x_{1k}, ..., x_{Tk}\}$ is equivalent to minimizing the total squared error (68) over the same quantities, i.e., equivalent to solving the system $\Gamma_k X(k) = b_k$. An important feature of the equivalent smoothing filter is that it uses synthetic measurements

(74) and synthetic covariance matrices (70)-(71), not the given measurements $\{z_{t\ell}\}$ and originally specified covariances $\{Q_{tk}\}$ and $\{R_{tk}\}$.

The synthetic measurements $\widetilde{z}_{tk}$ form a probabilistic centroid over the sensor cells C. To see this, let

$$\widetilde{z}_{tk\ell} = \frac{\int_{B_\ell(t)} \tau \; N(\tau;\; H_{tk}x'_{tk},\; R_{tk})\; d\tau}{\int_{B_\ell(t)} N(\tau;\; H_{tk}x'_{tk},\; R_{tk})\; d\tau}, \qquad 1 \le \ell \le S, \tag{76}$$

denote the cell-level centroids for target k at time t. Because $\widetilde{z}_{tk\ell}$ is the mean of a PDF whose support is confined to the cell $B_\ell(t)$, it follows that $\widetilde{z}_{tk\ell} \in B_\ell(t)$ if $B_\ell(t)$ is a convex set. It is easily verified that

$$\widetilde{z}_{tk} = \frac{\sum_{\ell=1}^{S} \left[\frac{\bar{z}_{t\ell}}{P_\ell(X'_t)} \int_{B_\ell(t)} N(\tau;\; H_{tk}x'_{tk},\; R_{tk})\; d\tau\right] \widetilde{z}_{tk\ell}}{\sum_{\ell=1}^{S} \left[\frac{\bar{z}_{t\ell}}{P_\ell(X'_t)} \int_{B_\ell(t)} N(\tau;\; H_{tk}x'_{tk},\; R_{tk})\; d\tau\right]} \tag{77}$$

is equivalent to the definition (74). The synthetic measurement $\widetilde{z}_{tk}$ is a convex combination of the cell-level centroids $\{\widetilde{z}_{tk1}, ..., \widetilde{z}_{tkS}\}$ in (77).

The cell-level centroids (76) may be a convenient point in the calculation to deal with some of the numerical dynamic range issues that may arise. For example, in the one-dimensional case when cells $B_\ell(t)$ are intervals, it is straightforward to prove the identity

$$\gamma = \frac{\int_a^b \tau \; N(\tau;\mu,\sigma^2)\; d\tau}{\int_a^b N(\tau;\mu,\sigma^2)\; d\tau} = \mu - \sigma^2 \left(\frac{N(b;\mu,\sigma^2) - N(a;\mu,\sigma^2)}{\int_a^b N(\tau;\mu,\sigma^2)\; d\tau} \right). \tag{78}$$

The denominator is a difference of error functions, so evaluating the cell-level centroids in the one-dimensional case seems both fast and accurate. However, when the mean μ is separated from the interval $[a,b]$ by many multiples of σ, numerical underflow may cause both numerator and denominator of the ratio on the right hand side of (78) to become zero. An alternative approach is to show by a change of variables that

$$\gamma = c + \frac{\int_{-\Delta}^{\Delta} \tau \; N(\tau + c;\mu,\sigma^2)\; d\tau}{\int_{-\Delta}^{\Delta} N(\tau + c;\mu,\sigma^2)\; d\tau},$$

where $[a,b] = [c-\Delta, c+\Delta]$. Expanding numerator and denominator in a Maclaurin series, substituting the identity

$$\frac{d^n}{d\tau^n} N(\tau;\mu,\sigma^2) = \frac{(-1)^n}{\sigma^n}\; N(\tau;\mu,\sigma^2)\; He_n\left(\frac{\tau-\mu}{\sigma}\right), \qquad n = 0, 1, 2, ...,$$

where $He_n(\cdot)$ is the Hermite polynomial of degree n, and canceling the common factor of $N(\tau;\mu,\sigma^2)$ gives a standardized (dimensionless) ratio in the form

$$\frac{\gamma-c}{\sigma} = -\left(\frac{\Delta}{\sigma}\right)^2 \left(\frac{\sum_{n=0}^{\infty} \frac{1}{(2n+3)(2n+1)!} \left(\frac{\Delta}{\sigma}\right)^{2n} He_{2n+1}\left(\frac{c-\mu}{\sigma}\right)}{\sum_{n=0}^{\infty} \frac{1}{(2n+1)(2n)!} \left(\frac{\Delta}{\sigma}\right)^{2n} He_{2n}\left(\frac{c-\mu}{\sigma}\right)} \right).$$

Rational approximations of desired accuracy are obtained by truncating in the series; however, asymptotic approximations are preferable for sufficiently large values of $\frac{c-\mu}{\sigma}$. Approximations derived by methods such as these may be especially useful in higher dimensional applications.

6.3 COVARIANCE ESTIMATES

Measurement and target covariance matrices are assumed unknown in this section, and an estimation algorithm is derived by the generalized EM (GEM) method. The E-step of the GEM method is the same as the E-step of the EM method. Consequently, the mixing proportions $\{\pi_{tk}\}$ are estimated in exactly the same way as discussed earlier.

The terms in the $Q^\sharp$ function not involving mixing proportions constitute a function of the form $Q^\sharp(X,\{Q_{tk}\},\{R_{tk}\})$. The M-step of the GEM method requires solving the problem

$$\max_{X,Q,R} \; Q^\sharp(X,\{Q_{tk}\},\{R_{tk}\}). \tag{79}$$

Unfortunately, the necessary equations obtained by differentiation with respect to the state and covariance variables are coupled, so the EM method is difficult to use. By replacing the maximization (79) with two nested maximizations

$$\max_{Q,R} \left\{ \max_{X} Q^\sharp(X,\{Q_{tk}\},\{R_{tk}\}) \right\}, \tag{80}$$

it is readily verified that the Q function is necessarily increased, even though it is not maximized. Any increase is sufficient to ensure that the convergence properties of the EM method apply here as well [3, 8]. The covariance matrix estimators obtained in this way using (80) are therefore GEM estimators rather than EM estimators.

Let $\{\widehat{x}_{0k},\widehat{x}_{1k},...,\widehat{x}_{Tk}\}$ denote the updated state estimates obtained as in the previous section for current estimates of mixing proportions, states, and covariances. The notation of the total squared error expression (68) must be adjusted slightly for the present context; that is, $N(\tau;H_{tk}x'_{tk},R_{tk})$ must be replaced by $N(\tau;H_{tk}x'_{tk},R'_{tk})$, as is easily seen from the derivation of Q_{kX}, and x_{tk} must be replaced by $\widehat{x}_{tk}$ because of the nested maximization (80). Taking the gradient of the adjusted total error expression with respect to R_{tk} using the general matrix gradient identity

$$\nabla_\Sigma \log N(x;\mu,\Sigma) = -\Sigma^{-1} + \Sigma^{-1}(x-\mu)(x-\mu)^*\Sigma^{-1}$$

and solving for R_{tk} gives the estimator

$$\widehat{R}_{tk} = \frac{\sum_{\ell=1}^{S} \left[\frac{\overline{z}_{t\ell}}{P_\ell(X_t')} \int_{B_\ell(t)} N(\tau; H_{tk}x_{tk}', R_{tk}')\, d\tau \right] \widehat{R}_{tk\ell}}{\sum_{\ell=1}^{S} \left[\frac{\overline{z}_{t\ell}}{P_\ell(X_t')} \int_{B_\ell(t)} N(\tau; H_{tk}x_{tk}', R_{tk}')\, d\tau \right]}, \tag{81}$$

where the cell-level measurement covariance matrix contributions are defined by

$$\widehat{R}_{tk\ell} = \frac{\int_{B_\ell(t)} N(\tau; H_{tk}x_{tk}', R_k')\, (\tau - H_{tk}\widehat{x}_{tk})\, (\tau - H_{tk}\widehat{x}_{tk})^*\, d\tau}{\int_{B_\ell(t)} N(\tau; H_{tk}x_{tk}', R_{tk}')\, d\tau}.$$

The estimator (81) cannot be full rank unless $S > \dim(C)$. Also, numerical considerations analogous to those mentioned at the end of the previous subsection apply to computing $\widehat{R}_{tk\ell}$. Further details are not pursued here.

If all the measurement covariances are required to be the same, so that $R_{tk} \equiv R$ for all t, then the same procedure gives

$$\widehat{R}_k = \frac{\sum_{t=1}^{T} \pi_{tk}' \sum_{\ell=1}^{S} \left[\frac{\overline{z}_{t\ell}}{P_\ell(X_t')} \int_{B_\ell(t)} N(\tau; H_{tk}x_{tk}', R_{tk}')\, d\tau \right] \widehat{R}_{tk\ell}}{\sum_{t=1}^{T} \pi_{tk}' \sum_{\ell=1}^{S} \left[\frac{\overline{z}_{t\ell}}{P_\ell(X_t')} \int_{B_\ell(t)} N(\tau; H_{tk}x_{tk}', R_k')\, d\tau \right]}. \tag{82}$$

The estimator (82) cannot be full rank unless $ST > \dim(C)$. Both estimators (81) and (82) are convex (discrete-continuous) combinations of outer products of innovations.

A similar method leads to target covariance matrix estimates. Taking the gradient of the total squared error (68) with respect to $Q_{t-1,k}$ and solving for $Q_{t-1,k}$ gives

$$\widehat{Q}_{t-1,k} = (\widehat{x}_{tk} - F_{t-1,k}\widehat{x}_{t-1,k})\, (\widehat{x}_{tk} - F_{t-1,k}\widehat{x}_{t-1,k})^*. \tag{83}$$

The estimator (83) is full rank only if $\dim(x_{tk}) = 1$. When $\dim(x_{tk}) > 1$, taking the gradient of the total squared error under the additional constraints $Q_{tk} \equiv Q_k$ for all t gives

$$\widehat{Q}_k = \frac{\sum_{t=1}^{T} \frac{\|Z_t\|}{P(X_t')} (\widehat{x}_{tk} - F_{t-1,k}\widehat{x}_{t-1,k})\, (\widehat{x}_{tk} - F_{t-1,k}\widehat{x}_{t-1,k})^*}{\sum_{t=1}^{T} \frac{\|Z_t\|}{P(X_t')}}. \tag{84}$$

The estimator (84) cannot be full rank unless $T > \dim(x_{tk})$.

Additional constraints can be imposed on the problem if desired. For example, if the k-th target and measurement processes are stationary, then $R_{tk} \equiv R$ and $Q_{tk} \equiv Q$ for all t. Linear constraints such as this lead to estimators that average over both ℓ and t, so the resulting covariance estimators are more likely to have full rank. Alternatively, Wishart matrix prior PDFs can be utilized to guarantee that the estimated covariance matrices have full rank [10]. Such variations are application dependent and are not pursued here.

7. STATEMENT OF THE HISTOGRAM-PMHT ALGORITHM

Many application-specific variants of the histogram-PMHT algorithm are easily derived from the above analysis. The assumptions of the specific algorithm given in this section are now reiterated. The sensor cells C are known and of fixed size and shape. The numbers $\{L(1), ..., L(T)\}$ of displayed cells are given, and the displayed cell list $B(t) = \{B_1(t), ..., B_{L(T)}(t)\}$ is known for $t = 1, ..., T$. The measurement vectors $\{Z_1, ..., Z_T\}$ are in the form specified by equation (1).

The target covariance matrices are stationary; that is, only one covariance matrix is estimated per target. The measurement covariance matrices are non-stationary, so a covariance matrix is estimated for each target at each time. The T background (noise) PDFs

$$\{G_0(\tau; \overline{x}_{10}), \ ..., \ G_0(\tau; \overline{x}_{T0})\}$$

and their corresponding parameters $\{\overline{x}_{10}, \overline{x}_{20}, ..., \overline{x}_{T0}\}$ are given. For computational efficiency, the numerical values of the quantities $\{\|Z_1\|, ..., \|Z_T\|\}$ and the integrals

$$\int_{B_\ell(t)} G_0(\tau; \ \overline{x}_{t0}) \, d\tau, \quad \ell = 1, ..., S \ \text{ and } \ t = 1, ..., T,$$

are precomputed.

All target models are assumed to be observable; that is, a unique point estimate exists for each target state given the measured data. This assumption may impose mild restrictions in some applications. In some kinematic applications, for example, measurements of position are used to estimate a target state vector comprising both position and velocity. In this case, it is clearly necessary to have at least two scans in the batch, unless an informative *a priori* state density is available. For the single batch histogram-PMHT algorithm, *a priori* information is absent, so in this case it is required that $T \geq 2$. In contrast, a single scan can be used in the recursive histogram-PMHT batch algorithm because it uses an informative *a priori* density obtained via earlier measured data; thus, the single batch algorithm is essentially an initialization procedure for the recursive batch algorithm. General conditions for observability of the target state are not pursued further here.

7.1 SINGLE BATCH HISTOGRAM-PMHT ALGORITHM

7.1.1 *Initialization*

Initialize mixing proportions $\{\widehat{\pi}_{tk}^{(0)}\}$ so that $\widehat{\pi}_{tk}^{(0)} > 0$ and

$$\widehat{\pi}_{t0}^{(0)} + \widehat{\pi}_{t1}^{(0)} + ... + \widehat{\pi}_{tM}^{(0)} = 1 \qquad \text{for } \ t = 1, ..., T.$$

For target models $k = 1, ..., M$, initialize

target state sequence:	$\{\widehat{x}_{0k}^{(0)}, \widehat{x}_{1k}^{(0)}, ..., \widehat{x}_{Tk}^{(0)}\}$,
measurement covariance sequence:	$\{\widehat{R}_{1k}^{(0)}, \widehat{R}_{2k}^{(0)}, ..., \widehat{R}_{Tk}^{(0)}\}$,
target covariance:	$\widehat{Q}_k^{(0)}$.

The initial covariance matrices are assumed to be valid covariance matrices; that is, they must be symmetric and positive-definite.

7.1.2 *Iteration*

Let i denote the histogram-PMHT iteration index. For $i = 0, 1, 2, ...$, compute the following 13 quantities:

1. Target cell probabilities for $t = 1, ..., T$; $\ell = 1, ..., S$; and $k = 0, 1, ..., M$:

$$P_{tk\ell}^{(i+1)} = \begin{cases} \int_{B_\ell(t)} G_0(\tau;\ \overline{x}_{t0})\ d\tau, & k = 0, \\ \int_{B_\ell(t)} N\left(\tau;\ H_{tk}\,\widehat{x}_{tk}^{(i)},\ \widehat{R}_{tk}^{(i)}\right)\ d\tau, & k = 1, ..., M. \end{cases}$$

2. Total cell probabilities for $t = 1, ..., T$ and $\ell = 1, ..., S$:

$$P_{t\ell}^{(i+1)} = \sum_{k=0}^{M} \widehat{\pi}_{tk}^{(i)} P_{tk\ell}^{(i+1)}. \tag{85}$$

3. Total sensor probabilities for $t = 1, ..., T$:

$$P_t^{(i+1)} = \sum_{\ell=1}^{L(t)} P_{t\ell}^{(i+1)}. \tag{86}$$

4. Expected measurements for $t = 1, ..., T$ and $\ell = 1, ..., S$:

$$\overline{z}_{t\ell}^{(i+1)} = \begin{cases} z_{t\ell}, & 1 \le \ell \le L(t) \\ \|Z_t\| \left(P_{t\ell}^{(i+1)} \Big/ P_t^{(i+1)}\right), & L(t)+1 \le \ell \le S. \end{cases}$$

5. Cell-level centroids for $t = 1, ..., T$; $\ell = 1, ..., S$; and $k = 1, ..., M$:

$$\widetilde{z}_{tk\ell}^{(i+1)} = \frac{1}{P_{tk\ell}^{(i+1)}} \int_{B_\ell(t)} \tau\ N\left(\tau;\ H_{tk}\,\widehat{x}_{tk}^{(i)},\ \widehat{R}_{tk}^{(i)}\right)\ d\tau.$$

6. Synthetic measurements for $t = 1, ..., T$ and $k = 1, ..., M$:

$$\widetilde{z}_{tk}^{(i+1)} = \frac{\sum_{\ell=1}^{S} \left[\overline{z}_{t\ell}^{(i+1)} \left(P_{tk\ell}^{(i+1)} \Big/ P_{t\ell}^{(i+1)}\right)\right] \widetilde{z}_{tk\ell}^{(i+1)}}{\sum_{\ell=1}^{S} \left[\overline{z}_{t\ell}^{(i+1)} \left(P_{tk\ell}^{(i+1)} \Big/ P_{t\ell}^{(i+1)}\right)\right]}.$$

7. Synthetic measurement covariance matrices for $t = 1, ..., T$ and $k = 1, ..., M$:

$$\widetilde{R}_{tk}^{(i+1)} = \frac{\widehat{R}_{tk}^{(i)}}{\widehat{\pi}_{tk}^{(i)} \sum_{\ell=1}^{S} \overline{z}_{t\ell}^{(i+1)} \left(P_{tk\ell}^{(i+1)} \Big/ P_{t\ell}^{(i+1)}\right)}.$$

8. Synthetic target covariance matrices for $t = 0, 1, ..., T-1$ and $k = 1, ..., M$:

$$\widetilde{Q}_{tk}^{(i+1)} = \frac{P_{t+1}^{(i+1)}}{\|Z_{t+1}\|} \widehat{Q}_k^{(i)}.$$

9. Estimated mixing proportions for $t = 1, ..., T$ and $k = 0, 1, ..., M$:

$$\widehat{\pi}_{tk}^{(i+1)} = \frac{\widehat{\pi}_{tk}^{(i)} \sum_{\ell=1}^{S} \overline{z}_{t\ell}^{(i+1)} \left(P_{tk\ell}^{(i+1)} \Big/ P_{t\ell}^{(i+1)} \right)}{\sum_{k'=0}^{M} \widehat{\pi}_{tk'}^{(i)} \sum_{\ell=1}^{S} \overline{z}_{t\ell}^{(i+1)} \left(P_{tk'\ell}^{(i+1)} \Big/ P_{t\ell}^{(i+1)} \right)}.$$

10. Estimated target states for $t = 0, 1, ..., T$ and $k = 1, ..., M$, using (for computational efficiency) a recursive Kalman smoothing filter, which comprises a forward filter initialized at time 0 with

$$\widetilde{y}_{0|0}^{(i+1)}(k) = 0, \tag{87}$$

and the dummy covariance matrix

$$P_{0|0}^{(i+1)}(k) = 0, \tag{88}$$

and given, for $t = 0, 1, ..., T-1$, by the recursions

$$\begin{aligned}
P_{t+1|t}^{(i+1)}(k) &= F_{tk}\, P_{t|t}^{(i+1)}(k)\, F_{tk}^{*} + \widetilde{Q}_{tk}^{(i+1)}, \\
W_{t+1}^{(i+1)}(k) &= P_{t+1|t}^{(i+1)}(k)\, H_{t+1,k}^{*} \left\{ H_{t+1,k}\, P_{t+1|t}^{(i+1)}(k)\, H_{t+1,k}^{*} + \widetilde{R}_{t+1,k}^{(i+1)} \right\}^{-1}, \\
P_{t+1|t+1}^{(i+1)}(k) &= P_{t+1|t}^{(i+1)}(k) - W_{t+1}^{(i+1)}(k)\, H_{t+1,k}\, P_{t+1|t}^{(i+1)}(k), \\
\widetilde{y}_{t+1|t+1}^{(i+1)}(k) &= F_{tk}\, \widetilde{y}_{t|t}^{(i+1)}(k) + W_{t+1}^{(i+1)}(k) \left\{ \widetilde{z}_{t+1,k}^{(i+1)} - H_{t+1,k}\, F_{tk}\, \widetilde{y}_{t|t}^{(i+1)}(k) \right\},
\end{aligned}$$

and a backward filter initialized at time T with

$$\widehat{x}_{Tk}^{(i+1)} \equiv \widetilde{y}_{T|T}^{(i+1)}(k)$$

and given, for $t = T-1, ..., 2, 1$, by the recursion

$$\widehat{x}_{tk}^{(i+1)} = \widetilde{y}_{t|t}^{(i+1)}(k) + P_{t|t}^{(i+1)}(k)\, F_{tk}^{*} \left(P_{t+1|t}^{(i+1)}(k) \right)^{-1} \left\{ \widehat{x}_{t+1,k}^{(i+1)} - F_{tk}\, \widetilde{y}_{t|t}^{(i+1)}(k) \right\} \tag{89}$$

and, for $t = 0$, by

$$\widehat{x}_{0k}^{(i+1)} = \left[F_{0k}^{*} \left(\widetilde{Q}_{0k} \right)^{-1} F_{0k} \right]^{-1} F_{0k}^{*} \left(\widetilde{Q}_{0k} \right)^{-1} \widehat{x}_{1k}^{(i+1)}. \tag{90}$$

11. Cell-level measurement covariance contributions for $t = 1, ..., T$; $\ell = 1, ..., S$; and $k = 1, ..., M$:

$$\widehat{R}_{tk\ell}^{(i+1)} = \frac{\int_{B_\ell(t)} N\left(\tau;\, H_{tk}\, \widehat{x}_{tk}^{(i)}, \widehat{R}_{tk}^{(i)} \right) \left(\tau - H_{tk}\, \widehat{x}_{tk}^{(i+1)} \right) \left(\tau - H_{tk}\, \widehat{x}_{tk}^{(i+1)} \right)^{*} d\tau}{P_{tk\ell}^{(i+1)}}.$$

12. Estimated measurement covariance matrices for $t = 1, ..., T$ and $k = 1, ..., M$:

$$\widehat{R}_{tk}^{(i+1)} = \frac{\sum_{\ell=1}^{S} \left[\overline{z}_{t\ell}^{(i+1)} \left(P_{tk\ell}^{(i+1)} \Big/ P_{t\ell}^{(i+1)} \right) \right] \widehat{R}_{tk\ell}^{(i+1)}}{\sum_{\ell=1}^{S} \left[\overline{z}_{t\ell}^{(i+1)} \left(P_{tk\ell}^{(i+1)} \Big/ P_{t\ell}^{(i+1)} \right) \right]}.$$

13. Estimated target covariance matrices for $k = 1, ..., M$:

$$\widehat{Q}_{k}^{(i+1)} = \frac{\sum_{t=1}^{T} \frac{\|Z_t\|}{P_t^{(i+1)}} \left(\widehat{x}_{tk}^{(i+1)} - F_{t-1,k}\, \widehat{x}_{t-1,k}^{(i+1)} \right) \left(\widehat{x}_{tk}^{(i+1)} - F_{t-1,k}\, \widehat{x}_{t-1,k}^{(i+1)} \right)^{*}}{\sum_{t=1}^{T} \left(\|Z_t\| \Big/ P_t^{(i+1)} \right)}.$$

This completes one step of the histogram-PMHT algorithm for a single batch.

The only theoretically justified convergence tests are based on the rate of increase in the overall likelihood function. In practice, effective tests can be based on the rate of change of the target state estimates or on other suitable grounds. Another strategy that may also be effective in practice is simply to compute a fixed number of iterations and terminate. However, useful alternative convergence tests are application-dependent and are not pursued here.

During algorithm development, it is highly recommended that the overall likelihood function be calculated and its monotonic increase verified. Monotonicity is an acid test for validating both software implementations and numerical dynamic range handling procedures; the availability of such a simple test is fortunate considering the detailed structure of the histogram-PMHT algorithm.

Upon convergence at, say, EM iteration i^*, the last estimates obtained are renamed as follows, for $1 \leq k \leq M$:

$$\begin{aligned}
\widehat{x}_{t|T}(k) &= \widehat{x}_{tk}^{(i^*)} \quad \text{for } 0 \leq t \leq T, \\
\widehat{\pi}_{t|T}(k) &= \pi_{tk}^{(i^*)} \quad \text{for } 1 \leq t \leq T, \\
\widehat{R}_{t|T}(k) &= \widehat{R}_{tk}^{(i^*)} \quad \text{for } 1 \leq t \leq T, \\
\widehat{Q}_{t|T}(k) &= \widehat{Q}_{k}^{(i^*)} \quad \text{for } 0 \leq t \leq T-1.
\end{aligned} \tag{91}$$

The traditional notation of smoothing filters is adopted here to clarify the relationships between the various parameter estimates in the recursive batch application.

This completes the statement of the histogram-PMHT algorithm for a single batch using all the available data.

Error covariance matrices for the state estimates $\widehat{x}_{t|T}(k)$ are not required by the single batch histogram-PMHT algorithm, but they are needed in the recursive batch algorithm. The error covariance matrix of $\widehat{x}_{T|T}(k)$ is given by

$$\widehat{\Sigma}_{T|T}(k) = P_{T|T}^{(i^*)}(k).$$

For $t = T-1, ..., 1, 0$, the error covariance matrix of $\widehat{x}_{t|T}(k)$ is given by the backward recursion

$$\begin{aligned}\widehat{\Sigma}_{t|T}(k) &= P_{t|t}^{(i^*)}(k) + P_{t|t}^{(i^*)}(k)\, F_{tk}^{*} \left(P_{t+1|t}^{(i^*)}(k)\right)^{-1} \\ &\quad \times \left[\widehat{\Sigma}_{t+1|T}(k) - P_{t+1|t}^{(i^*)}(k)\right] \left(P_{t+1|t}^{(i^*)}(k)\right)^{-1} F_{tk}\, P_{t|t}^{(i^*)}(k), \end{aligned} \tag{92}$$

where the quantities $P_{t|t}^{(i^*)}(k)$ and $P_{t+1|t}^{(i^*)}(k)$ are the intermediate covariance matrices computed in step 10 above, after algorithm convergence.

7.2 RECURSIVE BATCH ALGORITHM

It is assumed in this section that data vectors Z_t are available for scans at times $t = 1, 2, 3, ...$, and that scans are processed by the histogram-PMHT algorithm using a sliding batch of maximum length T. In the startup phase when current time t is such that $1 \leq t \leq T$, the histogram-PMHT algorithm is applied to all available scans. The batch length thus grows steadily and is always equal to t. The batch is full for the first time when $t = T$. Thereafter, the batch is refreshed by adjoining the newest scan and deleting the oldest scan, thereby keeping batch length T fixed.

Let $t \geq 1$ denote the time of the most recent scan. During the startup phase, the histogram-PMHT algorithm described in the preceding section is used without modification for the current batch length. Using the notation in (91) and (92), the histogram-PMHT outputs at time t are now denoted as

$$\left\{\widehat{x}_{n|t}(k),\ \widehat{\Sigma}_{n|t}(k),\ \widehat{\pi}_{n|t}(k),\ \widehat{R}_{n|t}(k),\ \widehat{Q}_{n|t}(k)\right\} \quad \text{for} \quad n = 0, 1, ..., t.$$

This completes the startup phase.

When current time $t > T$, it is necessary to remove the oldest data scan from the batch. The current batch comprises data from the most recent T scans,

$$\{Z_{t-T+1},\ Z_{t-T+2},\ ...,\ Z_{t-1},\ Z_t\}.$$

The set of all deleted data vectors is

$$\{Z_1, Z_2, ..., Z_{t-T}\}.$$

Let $p_{\Xi_{t-T}}(X_{t-T} \mid N_1, ..., N_{t-T})$ denote the posterior PDF of X_{t-T} conditioned on the quantized vectors $N_1, ..., N_{t-T}$ that correspond to the deleted data vectors $\{Z_1, Z_2, ..., Z_{t-T}\}$. Adjusting the notation in the obvious way to accommodate the sliding batch gives the modified data-dependent *a priori* density as

$$\left[p_{\Xi_{t-T}}(X_{t-T} \mid N_1, ..., N_{t-T})\right]^{N_{t-T,\Sigma}+N_{t-T,\Sigma}^{c}} \prod_{n=t-T+1}^{t} \left[p_{\Xi_n|\Xi_{n-1}}(X_n|X_{n-1})\right]^{N_{n,\Sigma}+N_{n,\Sigma}^{c}}. \tag{93}$$

The only significant difference between (29) and (93) is that the exponent of the leading term in (93) is changed because the data at scan $t-T$ are generated in the manner described in section 3.3. The auxiliary function $Q^{\sharp}(t)$ is found by taking the limit as $\hbar^2 \to 0$ (*cf.* Section 4.4). The result is

$$\begin{aligned}
&\frac{\|Z_{t-T}\|}{P(X'_{t-T})} \log p_{\Xi_{t-T}}(X_{t-T}|Z_1, ..., Z_{t-T}) \\
&+ \sum_{n=t-T+1}^{t} \frac{\|Z_n\|}{P(X'_n)} \log p_{\Xi_n|\Xi_{n-1}}(X_n|X_{n-1}) \\
&+ \sum_{k=0}^{M} \sum_{n=n-T+1}^{t} \sum_{\ell=1}^{S} \frac{\overline{z}_{n\ell}}{P_\ell(X'_n)} \int_{B_\ell(n)} f_k(\tau|X'_n) \, \log f_k(\tau|X_n) \, d\tau.
\end{aligned} \tag{94}$$

The primary difference between (94) and (56) is the additional term contributed at time $t-T$. The M state estimates $\{\widehat{x}_{t-T|t-T}(k)\}$ are fixed and not subject to further update, so it is assumed that

$$p_{\Xi_{t-T}}(X_{t-T} \mid N_1, ..., N_{t-T}) = \prod_{k=1}^{M} N\left(x_{t-T,k};\, \widehat{x}_{t-T|t-T}(k),\, \widehat{\Sigma}_{t-T|t-T}(k)\right). \tag{95}$$

Substituting the approximation (95) gives the modified form of the total squared error (68) of the k-th target. The additional term at time $t-T$ gives an additional term in the upper left-hand corner block of the matrix of the necessary equations, denoted by $\Gamma_k X(k) = b_k$ in section 6.2. The equivalent Kalman smoothing filter (*cf.* equation (75)) now has a diffuse *a priori* density, as is seen from the equivalent likelihood expression

$$\begin{aligned}
&N\left(x_{t-T,k};\, \widehat{x}_{t-T|t-T}(k),\, \frac{P(X'_{t-T})}{\|Z_{t-T}\|}\widehat{\Sigma}_{t-T|t-T}(k)\right) \\
&\quad \times \prod_{n=t-T+1}^{t} \left\{N\left(x_{n,k};\, F_{n-1,k}\, x_{n-1,k},\, \widetilde{Q}_{n-1,k}\right)\, N\left(\widetilde{z}_{nk};\, H_{nk}\, x_{nk},\, \widetilde{R}_{nk}\right)\right\}.
\end{aligned}$$

Further details are straightforward and have been omitted.

For $t > T$, after the batch is refreshed, the histogram-PMHT algorithm described in section 7.1 is applied to the data set

$$\{Z_{t-T+1}, Z_{t-T+2}, ..., Z_{t-1}, Z_t\}$$

to compute the parameter estimates. However, the initializations (87) and (88) of the Kalman smoothing filter are changed to

$$\widetilde{y}^{(i+1)}_{0|0}(k) = \widehat{x}_{t-T|t-T}(k) \tag{96}$$

and the covariance matrix

$$P_{0|0}^{(i+1)}(k) = \frac{P_{t-T}^{(i+1)}}{\|Z_{t-T}\|}\widehat{\Sigma}_{t-T|t-T}(k), \tag{97}$$

respectively, where the probability $P_{t-T}^{(i+1)}$ is defined as in equation (86) using the current algorithm iterates $\widehat{X}_{t-T|t}^{(i)}(k)$ and $\widehat{R}_{t-T|t}^{(i)}(k)$. Also, the backward recursion (89) now runs from $t = T-1$ to $t = 0$ because the special step (90) is not required when the *a priori* density is not diffuse.

The outputs of the histogram-PMHT algorithm, including the error covariance matrices of the state estimates, are denoted in the smoothing filter notation as

$$\begin{array}{ll}
\widehat{x}_{n|t}(k) & \text{for } t-T \leq n \leq t, \\
\widehat{\Sigma}_{n|t}(k) & \text{for } t-T \leq n \leq t, \\
\widehat{\pi}_{n|t}(k) & \text{for } t-T+1 \leq n \leq t, \\
\widehat{R}_{n|t}(k) & \text{for } t-T+1 \leq n \leq t, \\
\widehat{Q}_{n|t}(k) & \text{for } t-T \leq n \leq t-1.
\end{array}$$

This completes the statement of the recursive form of the histogram-PMHT algorithm.

8. VARIATIONS ON THE THEME

Independent target models have been assumed throughout the report to allow clear focus on the central ideas and to simplify the analysis; however, this assumption can be modified significantly without changing either the fundamental theoretical framework or the structure of the histogram-PMHT algorithm. For example, the assumption of independent targets is incorrect when the targets are models of frequency lines and these lines are known to be harmonically related. It is straightforward in this case to use the functional dependencies between the lines as linear equality constraints on the line parameters; that is, the means and covariances of harmonically related lines satisfy specified linear constraints (transformations). Linear equality constraints are readily incorporated into the M-step of the histogram-PMHT algorithm, as indicated in section 6. For further details of this particular application, see Luginbuhl [2, chapter 5].

The basic strategy for modeling dependencies between targets can be generalized significantly and adapted to other applications. Thus, known functional relationships arising from physical or other considerations in the application can be modeled as parametric constraints on the M-step optimization problem. The difficulty of solving the M-step will depend heavily on the mathematical character of the constraints and will, in general, require the use of numerical procedures. The practicality of the resulting algorithm will be application-dependent in general.

A single Gaussian density may be inadequate in some applications to capture the cell-to-cell variation of one or more of the targets on the display. Loosely speaking, such a target may be best described as a "blob." Such problems may arise, for example, when the underlying target signal model is poorly understood and not well matched to the signal processor whose outputs feed the sensor display. For such applications, it is probably worthwhile to use a mixture-of-mixtures density (see Streit and Luginbuhl [11]) to model the target's cell-to-cell variation. Specifically, if the k-th target is a blob, then the mixture model density (30) is such that component $G_k(\tau|X_t)$ is itself modeled as a Gaussian mixture whose parameters must be estimated. The derivation of the histogram-PMHT algorithm for this case requires adding a fourth stage of missing data, namely, the assignment of measured data to components within the blob mixture. While such an algorithm for blob-like targets is easily derived using the methods presented in this report, it is left for future work.

9. SUMMARY

The histogram-PMHT algorithm is a multi-target tracking algorithm designed to be used with the entire sensor output data stream. It completely avoids the thresholding losses incurred by the traditional methods of generating point measurements by peak picking, three-point interpolation, etc. The theoretical development of the histogram-PMHT algorithm presented in this report has a mathematically sound foundation based on the framework of PMHT.

The negative multinomial density is potentially very useful in applications in which data compression and thresholding procedures result in truncated sensor measurements for some sensor cells. In other applications, limited resource availability may require collecting measurements in only a subset of cells. The negative multinomial model compensates for missing measured data by using an expectation operator to extrapolate the given data into truncated cells; hence, it may reduce parameter estimation bias and other undesirable edge effects induced by cell truncation.

REFERENCES

[1] R. L. Streit and T. E. Luginbuhl, "Probabilistic Multi-Hypothesis Tracking," NUWC-NPT Technical Report 10,428, Naval Undersea Warfare Center Division, Newport, Rhode Island, 15 February 1995.

[2] T. E. Luginbuhl, "Estimation of General Discrete-Time FM Processes," Ph. D. Thesis, University of Connecticut, Storrs, 1999.

[3] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum Likelihood from Incomplete Data via the EM Algorithm (with Discussion)," *Journal of the Royal Statistical Society B*, vol. 39, 1977, pp. 1-38.

[4] A. P. Dempster and D. B. Rubin, "Rounding Error in Regression: The Appropriateness of Sheppard's Corrections," *Journal of the Royal Statistical Society B*, vol. 45, 1983, pp. 51-59.

[5] G. J. McLachlan and P. N. Jones, "Fitting Mixture Models to Grouped and Truncated Data via the EM Algorithm," *Biometrics*, vol. 44, 1988, pp. 571-578.

[6] P. N. Jones and G. J. McLachlan, "Algorithm AS 254. Maximum Likelihood Estimation from Grouped and Truncated Data with Finite Normal Mixture Models," *Applied Statistics*, vol. 39, 1990, pp. 273-282.

[7] P. N. Jones and G. J. McLachlan, "Improving the Convergence Rate of the EM Algorithm for a Mixture Model Fitted to Grouped Truncated Data," *Journal of Statistical Computation and Simulation*, vol. 43, 1992, pp. 31-44.

[8] G. J. McLachlan and T. Krishnan, *The EM Algorithm and Extensions*, Wiley & Sons, New York, 1997.

[9] N. L. Johnson, S. Kotz, and N. Balakrishnan, *Discrete Multivariate Distributions*, Wiley & Sons, New York, 1997.

[10] R. A. Redner, R. J. Hathaway, and J. C. Bezdek, "Estimating the Parameters of Mixture Models with Modal Estimators," *Communications in Statistics, Part A: Theory and Methods*, vol. 16, 1987, pp. 2639-2660.

[11] R. L. Streit and T. E. Luginbuhl, "Maximum Likelihood Training of Probabilistic Neural Networks," *IEEE Transactions on Neural Networks*, vol. NN-5, no. 4, 1994, pp. 764-783.

INITIAL DISTRIBUTION LIST

Addressee	No. of Copies
Naval Sea Systems Command (SEA 93 (ASTO D1)--R. Zarnich)	1
Defense Technical Information Center	2
MITRE Corp. (G. Jacyna)	1
Lockheed Martin Corp. (A. Cohen)	1
ORINCON, Inc. (G. Godshalk, D. Pace, T. Carson)	3
Engineering Technology Center, A&T, Inc. (L. Gorham, T. Parberry, K. Richards)	3
Metron, Inc. (L. Stone)	1
Digital System Resources, Inc. (J. Ambrose, B. Shapo)	2
AETC (K. Delanoy)	1
University of Connecticut (P. Willett)	1
University of Massachusetts Dartmouth (S. Nardone)	1